DOCUMENT RESUME

ED 189 131 TM 800 297

AUTHOR Peters, Eileen
TITLE Holistic Scoring Manual. Basic Skills Improvement Policy Implementation Guide #2.
INSTITUTION Massachusetts State Dept. of Education, Boston. Bureau of Research and Assessment.
PUB DATE Oct 79
NOTE 58p.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Achievement Rating; Basic Skills; Educational Testing; Elementary Secondary Education; *Essay Tests; Examiners; Guidelines; *Program Administration; Scheduling; *Scoring; Student Evaluation; *Testing Programs; Writing (Composition); *Writing Skills
IDENTIFIERS *Holistic Evaluation; Massachusetts

ABSTRACT

The technical and administrative processes and activities involved in scoring student responses to essay writing questions in the Massachusetts State Board of Education's basic skills program are outlined and discussed. After a general introduction to holistic scoring, separate sections describe the roles and activities of the administrators, the readers, the chief and assistant chief readers, and the aides. The section for the administrator addresses supervisory functions, delineates tasks that must be accomplished prior to the reading, suggests appropriate timelines, includes worksheets for estimating scoring time and the number of readers required. The section for the readers explains their responsibilities and notes the procedures that will be followed at the scoring session. The section for the chief reader and the assistant chief reader explains, in the order of occurence, the steps that should be followed in planning the reading, selecting training papers, training readers, and scoring the papers. Schedules and timelines are suggested, and quality control checklists are provided. The section for the aides details their responsibilities prior to and during the scoring session. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED189131

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Basic Skills Improvement Policy

Implementation Guide #2

HOLISTIC SCORING MANUAL

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
C. DiBella
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Massachusetts
Department of Education
October, 1979

Massachusetts Board of Education

Charles T. Grigsby, Chairperson
Ann H. McHugh, Vice Chairperson

John Anthony
Mary Barros
John W. Bond
James L. Green
Howard A. Greis
Mary Ann Hardenbergh
Joseph C. Mello, Jr.
Edwin M. Rossman

Gregory R. Anrig, Commissioner of Education, Secretary

Ex. Officio Member:

Laura Clausen, Chancellor, Board of Education

Bureau of Research and Assessment

Allan S. Hartman
Tracy Libros
Leslie S. May
Matthew H. Towle

The Massachusetts Department of Education insures equal employment/ educational opportunities/and affirmative action, regardless of race, color, creed, national origin, or sex, in compliance with Title IX, or handicap, in compliance with Section 504.

Publication #11568 approved by Alfred C. Holland, State Purchasing Agent

HOLISTIC SCORING MANUAL

by

Eileen Peters
New England Office
Educational Testing Service

for

Massachusetts Department of Education
Bureau of Research and Assessment
31 St. James Avenue
Boston, Massachusetts 02116

October, 1979

FEB 1 3 1980

ACKNOWLEDGMENTS

The guidelines discussed in this manual are based on Mary E. Fowles' description of holistic scoring procedures in Educational Testing Service's Manual for Scoring the Writing Sample for the Basic Skills Assessment.

I am most grateful to Estelle Bartels for contributing to the clarity of the manuscript, and to F. Terry Smith for overseeing its production.

Eileen Peters
New England Office
Educational Testing Service

CONTENTS

		page
1.	INTRODUCTION	
	What is the relationship of this manual to the Basic Skills Improvement Policy?	1
	What is holistic scoring?	2
	How are minimal standards established?	3
	How is this manual organized for your convenience?	4
2.	FOR THE ADMINISTRATOR	
	Why is the administrator's role critical?	6
	When should the scoring session planning begin?	7
	How much time should be allowed for the scoring session and how many readers will be needed?	7
	When should the scoring session be held?	8
	Do you need an Assistant Chief Reader?	10
	How many aides are needed?	10
	Who should read and score the student papers?	11
	What kind of room arrangements are necessary?	12
	How should minimal standards be set?	12
	Exercise: Planning the length of the scoring session and estimating the number of readers needed	14
	Administrative Checklist	16
	Several weeks before the scoring session	16
	Two weeks before the scoring session	16
	Two days before the scoring session	17
3.	FOR THE READERS	18
	The training session	19
	Discussing problems and unusual papers	21
	The scoring	22
4.	FOR THE CHIEF READER AND THE ASSISTANT CHIEF READER	23
	Several Weeks Before the Reading	24
	Selecting and scoring training papers	24
	What are training papers?	24
	When should training papers be selected?	24
	How many papers should you read?	25
	How many and what types of papers should you select?	25
	How should you score the training papers?	26
	How should you order the papers for presentation?	27
	How should you label the training papers?	27
	What final preparations are needed?	28
	Making room arrangements	29
	What is the most efficient way to arrange the room for the scoring session?	29
	Arrangement of tables at scoring session	31
	Scheduling the scoring session	32
	Sample agenda	33
	Planning the distribution system for the scoring session	34
	How should you organize the distribution of student papers to readers?	34
	Devising a scoring sheet	34

Why should you have a scoring sheet? 34
Example: Scoring Forms 35
Two Days Before the Reading 36
What should you do? 36
Conducting the Reading 36
Preliminary discussion with readers 36
Training the readers 37
Scoring the papers 42
The next exercise 43

5. FOR THE AIDES 44
Several weeks before the scoring session 44
The day before the scoring session 46
During the scoring session 46

Chief Reader's Quality Control Checklist 48
Definitions 50

1. INTRODUCTION

WHAT IS THE RELATIONSHIP OF THIS MANUAL TO THE BASIC SKILLS IMPROVEMENT POLICY?

The purpose of the Massachusetts State Board of Education's Basic Skills Improvement Policy is to improve students' reading, writing, speaking, listening, and mathematics skills. All policy provisions covering planning, assessment, and remediation are essential to the accomplishment of that goal. To date, assessment - especially the development and selection of tests and writing exercises - has been the primary concern of Massachusetts educators. But while assessment instruments are critical, the related issue of scoring student responses is no less integral to basic policy considerations. The point is simply that if skill improvement is to occur, deficiencies must be recognized; local school districts must make professionally sound decisions in identifying students who need instructional assistance.

In the area of writing, the policy requires that writing be assessed by means of a writing sample. If the writing samples are scored holistically, school officials who have proper expertise in holistic scoring techniques will avoid two types of major errors. They will not systematically fail to detect students who need help, nor will they repeatedly misclassify students who possess the skills in question.

This manual explains the holistic scoring process and explicates the procedures involved in conducting a scoring session. The purpose is to help educators assess the quality of student writing.

WHAT IS HOLISTIC SCORING?

Holistic scoring means judging a piece of writing for its total effect. Teachers of writing read papers quickly and evaluate their total impact. If the writing contains a great many errors in spelling, syntax, or organization, the reader's impression of its effectiveness will be negatively influenced. However, the discrete elements are not as important as the total expression of a student's ideas and opinions—that is, the overall quality of the response.

At the beginning of every scoring session, the Chief Reader (person conducting the session) trains the readers (teachers of writing) in the procedures of holistic scoring. The readers discuss the topic students wrote on and then evaluate a set of training papers selected by the Chief Reader to represent the range of writing produced in this administration. They compare the papers with each other, not to preconceived ideas of a perfect paper. They assign a score of 4 (best), 3, 2, or 1 to each paper and discuss any differences of opinion. Out of this discussion of training papers arise the guidelines or rationale for scoring the responses in this administration.

When the teachers have reached a consensus about how the papers are to be scored, they are ready to begin the actual scoring.

Each paper is judged by at least two different readers. If the two scores are the same (4 and 4, 3 and 3, 2 and 2, 1 and 1) or adjacent (4 and 3, 3 and 2, or 2 and 1), they are totaled to form the final score. For example, a paper that receives readers' scores of 3 and 2 has a final score of 5. If a paper receives two discrepant scores (4 and 2, 4 and 1,

or 3 and 1), it must be rescored by two different readers. If the two new scores are also discrepant, the Chief Reader and Assistant Chief Reader assign a final score to that paper.

The papers are anonymous; readers cannot tell who wrote the papers they are scoring. Also, readers record scores on separate forms so that they will not know, and thus be influenced by, the first score a paper receives.

The holistic scoring procedures as outlined in print can seem complicated but, if used properly, they result in an extremely efficient and, as nearly as possible, reliable scoring session.

HOW ARE MINIMAL STANDARDS ESTABLISHED?

A final score, for example 4, is not by itself a minimal standard. People must establish a point on the score scale, at or below which papers are to be regarded as not meeting minimal local standards. This point ("cutting score") can be established in different ways.

For example, the readers can set minimal standards for each exercise they score. Following this method, the readers identify papers that they consider minimally acceptable in content and expression. These papers are then used to establish the cutting score. Readers may agree, for example, to establish the following explanation of scores:

Score		
	8 (4 + 4)	Meets minimal standards
	7 (4 + 3)	
	6 (3 + 3)	
	5 (3 + 2)	
	4 (2 + 2)	Does not meet minimal standards
	3 (2 + 1)	
	2 (1 + 1)	

Papers are still judged for their overall effectiveness, but their effectiveness is evaluated in terms of a level of competence set by the readers at the scoring session.

Alternatively, the cutting score can be set after all the papers are scored. Using this method, a panel (comprised perhaps of writing teachers, administrators, parents, employers) is assembled. The panel is briefed on the purpose of the exercise, scoring procedures, and district policy considerations. Panel members then read and discuss a few representative papers at each score level and decide the point at which the cutting score should be established.

Both procedures for establishing minimal writing standards are explained in a companion standard setting manual.

HOW IS THIS MANUAL ORGANIZED FOR YOUR CONVENIENCE?

This manual contains detailed step-by-step guidelines for conducting a holistic scoring session.

To facilitate planning and permit easy duplication, the manual is organized by roles in the holistic scoring process.

Section 2, "For the Administrator," addresses supervisory functions, delineates tasks that must be accomplished prior to the reading, suggests appropriate timelines, and includes worksheets for estimating scoring time and the number of readers required.

Section 3, "For the Readers," explains readers' responsibilities and notes the procedures that will be followed at the scoring session.

Section 4, "For the Chief Reader and the Assistant Chief Reader," explains, in the order of occurence, the steps that should be followed in

planning the reading, selecting training papers, training readers, and scoring the papers. Schedules and timelines are suggested, and quality control checklists are provided.

Section 5, "For the Aides," details aides' responsibilities prior to and during the scoring session.

2. FOR THE ADMINISTRATOR

Planning the Holistic Scoring:

Summary of Major Administrative Tasks

1. Determine (on the basis of the number of readers available and the number of essays to be read) the length of the scoring session.

2. Schedule the scoring session, allowing enough lead time for preliminary preparations such as coding papers.

3. Select a Chief Reader and Assistant Chief Reader who will conduct the scoring session; select aides to assist the Chief Reader in preparing for the scoring and in conducting the reading; select readers who will do the scoring.

4. Work with the Chief Reader to make the necessary room arrangements.

5. Decide whether you want minimal standards to be set by readers during the scoring session or by a panel (comprised of writing teachers, administrators, parents, and others) after the scoring is completed.

WHY IS THE ADMINISTRATOR'S ROLE CRITICAL?

There are two keys to a successful reading: first, the personnel involved must be appropriate for the tasks assigned. The Chief Reader, for example, should have considerable experience with holistic scoring and command the respect of colleagues; readers should be experienced teachers of writing, and aides should be extremely well organized and efficient. Second, the scoring session must be planned thoroughly, allowing ample time for preliminary preparations. If personnel are

carefully selected and all logistical details attended to, the scoring session will run smoothly. Participants will find the experience rewarding, and the score results will be useful for setting minimal writing standards and planning instructional follow-up.

The administrative procedures discussed below are designed to help you plan a productive scoring session. Before you attend to these responsibilities, however, you should read the remainder of this manual, particularly the sections covering the Chief Reader's responsibilities.

WHEN SHOULD THE SCORING SESSION PLANNING BEGIN?

Ideally, the scoring session should be planned several weeks before the writing exercises are administered. The participants, particularly the Chief Reader, must have time to prepare carefully and thoughtfully for their roles in the session. Although lead time might vary according to the size of the school system and its experience with holistic scoring, it is safe to allow <u>two months</u> for preparation.

HOW MUCH TIME SHOULD BE ALLOWED FOR THE SCORING SESSION AND HOW MANY READERS WILL BE NEEDED?

The length of the scoring session will depend upon: the number of writing exercises (topics students respond to), the number of papers written on each topic, the length of the average paper, the number of qualified readers available, and the experience of the Chief Reader and the readers.

On the average, it takes a trained reader approximately 45 seconds to score a piece of writing that covers one page. So a reader should be able to score about 40 pages each half hour. Of course, you will need to allow time for training periods and coffee breaks.

Logically, the scoring session can be shortened by increasing the number of readers. Two factors must be kept in mind, however. First, the readers should be teachers of writing who have had experience with a range of student writing ability. Second, a large reading can become unwieldy. A trained Chief Reader usually can supervise from 8 to 12 readers comfortably. If the Chief Reader has conducted a number of scoring sessions, s/he might manage from 12 to 25 readers. Larger readings probably should not be attempted unless they are supervised by experts in holistic scoring.

The exercise at the end of this section is designed to help you plan the length of the scoring session and the number of readers needed.

WHEN SHOULD THE SCORING SESSION BE HELD?

The ideal scoring session begins in the morning when readers are fresh and extends through the school day with sufficient time for discussion and breaks.

If it is necessary to schedule two or more half-day sessions, each must be planned so that <u>all</u> the student papers pertaining to a given topic are read on the same day. A split reading, with some papers read on one

day and the remainder read on another, will reduce score reliability appreciably.

Late-afternoon sessions, taking place at the close of the school day, should be avoided if at all possible. Experience shows that as readers tire, they score more slowly and less reliably. Moreover, their training will not be as productive or professionally rewarding.

WHO SHOULD BE APPOINTED CHIEF READER?

The Chief Reader conducts the holistic scoring sessions and is responsible for maintaining the scoring procedures outlined in this manual. Obviously, it is advantageous to have a Chief Reader who is experienced in the holistic scoring process. But if you decide to use someone untrained in holistic scoring, s/he must be a teacher of writing who is highly respected by his or her colleagues, who has had experience with a range of student writing ability, who is able to work firmly and fairly with others, and who fully understands the responsibilities of the Chief Reader.

Before selecting the Chief Reader, you should read the section of this manual that discusses the Chief Reader's responsibilities before and during the scoring session.

Regardless of whom you appoint, be certain that s/he has adequate preparation time. Three or four days (depending upon the complexity of the reading) will be needed to:

- Read this manual thoroughly
- Plan a schedule or agenda for the session
- Select training papers
- Train aides
- Supervise the preparation of student papers for scoring
- Check all arrangements carefully

DO YOU NEED AN ASSISTANT CHIEF READER?

The Assistant Chief Reader helps the Chief Reader select training papers and resolve score discrepancies; these decisions should not be made by one person. Thus, it is essential, even in holistic scoring sessions with only seven or eight readers, that the Chief Reader have an assistant. The Assistant Chief Reader should be a teacher of writing who has the respect of his or her colleagues and has had experience with a range of student writing ability. This may be considered a training position for future Chief Readers.

Ample preparation time, one or two days, must be allotted for the Assistant Chief Reader to:

- Help select training papers
- Read this manual carefully
- Plan the scoring sessions with the Chief Reader

HOW MANY AIDES ARE NEEDED?

Aides prepare materials before the holistic scoring sessions. At the scoring sessions, they distribute student papers and they record scores. Their work is supervised by the Chief Reader.

Aides may be parents, staff members, or even students who have not been tested. They must be efficient — both quick and thorough — and be willing to attend conscientiously to their specific instructions.

The number of aides also depends upon the size of the reading. For eight to twelve readers at a holistic reading, you will probably need three aides — two to distribute papers and one to check scores. For smaller readings, two aides may be sufficient.

You will probably want to designate one person as Chief Aide.

WHO SHOULD READ AND SCORE THE STUDENT PAPERS?

Readers for holistic scoring should be <u>teachers of writing</u>. Teachers not specifically trained to evaluate this skill tend to focus on one or two elements of writing to the exclusion of other competencies that must be assessed. As a consequence, they are less able to judge a piece of writing for its total effect which is a fundamental requirement for holistic scoring.

Not only should the readers be experienced teachers of writing, they also should have had experience with a broad range of student writing ability. Those who teach only Advanced Placement English, for example, or teachers who only provide remedial instruction would be inappropriate choices.

Similarly, readers' range of experience should encompass the appropriate grade levels, although some flexibility here is possible. If papers written by ninth grade students are to be scored, for example, readers might be selected from among sixth grade to twelth grade English teachers. Or, if seventh grade is the population of interest, readers could be drawn from grades four through ten.

Unlike those conducting the scoring session, readers will not need special preparation time. You will want to communicate with them beforehand, however, explaining why the scoring session is taking place and why you have chosen them to participate. Whether you hold a meeting or cover these points by memorandum, it would be helpful to give each reader a copy of "For the Readers", Section 3 of this manual.

WHAT KIND OF ROOM ARRANGEMENTS ARE NECESSARY?

The room in which the scoring session is held should be well-lighted, large, and comfortable. It should have:

- A large chalkboard.
- Tables to seat readers without crowding them (ideally four readers to a table).
- A head table for the Chief Reader and Assistant Chief Reader.
- Tables for the aides who will distribute papers and check scores.
- Sufficient room for the aides and Chief Reader to pass quickly among tables without disturbing the readers.

Diagrams of possible room arrangements are shown on page 31. You will want to discuss them with the Chief Reader before making final decisions.

HOW SHOULD MINIMAL STANDARDS BE SET?

Minimal standards can be set either during the reading or after the scoring is finished.

If they are established during the scoring, readers identify

papers that they consider minimally acceptable in content and expression. These papers are then used to establish a point on the scoring scale. Using a four-point scale of 4 (best) to 1, readers may agree, for example, to establish the following explanation of scores:

Score

4 - Is superior

3 - Meets minimal local standards

2 - Does not meet minimal local standards

1 - Is extremely weak

Papers will still be judged for their overall effectiveness, but their effectiveness will be evaluated in terms of a level of competence set by the readers.

Alternatively, papers can be compared with each other and scored on a scale of 4 (best) to 1 (worst) without regard to minimally acceptable scores. Since each paper is read twice final scores will range from 8 (4+4) to 2 (1+1). Then, after the scoring session, a panel can be assembled to examine the score distributions and sample papers, and set the minimal standard or cutting score.

You will want to read the section on setting writing standards in the standard setting manual before deciding when to set standards. If standard setting is to occur at the scoring session, this matter should be discussed thoroughly with the Chief Reader.

Exercise

PLANNING THE LENGTH OF THE SCORING SESSION AND ESTIMATING THE NUMBER OF READERS NEEDED

1. How many unique topics or separate writing assignments did students complete? _____ topics

2. Complete the following chart.

Number of papers and readings

Topic	Number of Papers	
1	_____	
2	_____	
3	_____	
Total	_____	(a)
Number of readings (a x 2)	_____	(b)

3. Assuming a reader can score one paper in 45 seconds, 80 papers in one hour, how many hours of actual reading time are needed $\frac{b}{(80)}$?

_____ hours

Example:

Total papers (a) 600

Number of readings (b) 1200

Hours $\frac{1200}{80}$ = 15 hours

4. Look at the sample schedule on page 33 of this manual. Allowing for necessary training and breaks, approximately 2 1/2 hours are available for actual scoring during the morning, and 1 1/2 hours

during the afternoon.

Allowing for training on <u>each</u> topic and breaks, how much actual reading time do you have available? _____ (c)

How many hours of reading time do you need? (from question 3) _____ (d)

How many readers are needed $\left(\frac{d}{c}\right)$? _____ Readers

Example: 2 1/2 hours (morning session) are available

15 hours are needed (previous example)

$\frac{15}{2.5}$ = 6 readers

(If 1 1/2 hours are available, $\frac{15}{1.5}$, or 10 readers, will be needed to score 600 papers.)

ADMINISTRATIVE CHECKLIST

several weeks before the scoring session

1. Plan the length of the scoring session, and determine the number of readers needed; schedule the session.

2. Select the Chief Reader, Assistant Chief Reader, aides, and readers.

3. Decide when standards are to be set.

4. Meet with the Chief Reader and Assistant Chief Reader to discuss holistic scoring procedures, and plan an agenda.

5. Communicate with the readers, explaining the purpose of the scoring session and the role they will play.

6. Make room arrangements.

7. Make arrangements for released time, substitutes, and other personnel matters; remember to provide sufficient preparation time.

two weeks before the scoring session

1. Meet with the Chief Reader and Assistant Chief Reader to review the agenda; review the progress they have made in planning the reading.

two days before the scoring session

1. Check room arrangements.

2. Check with the Chief Reader to be sure that training papers have been selected, coded, and duplicated, and that student papers are organized and ready to be scored.

3. Double check classroom coverage for readers, availability of aides, and other similar considerations.

3. FOR THE READERS

You are to be part of a holistic scoring process that assures that samples of student writing are judged as objectively as possible. You were selected to participate because you are experienced in dealing with a range of student writing.

When you use the holistic scoring method, you read a student's writing quickly and judge it for the total impact it makes on you. Even if you have never scored holistically, you will find that your experience with student writing enables you to make the quick and definite overall judgements required for holistic scoring. If a test paper contains a great many errors in spelling, syntax, or organization, those weaknesses will, of course, constitute part of your impression of the piece of writing. But what is more important in these exercises is the way the student has expressed his or her ideas and opinions——the overall quality of the writing.

Even the best papers will not be flawless. When you consider the constraints we impose upon the students —— the time limitations, the tenseness of the testing situation, you will realize why you and the other readers must not expect perfection or judge the writing sample against a preconceived "ideal" response. Instead, you should judge each test paper against the other test papers written for that exercise in this same test administration. The purpose is to identify which students have the greatest need of further writing instruction. To identify these students, you must compare their test papers only with each other.

In order to judge the test papers in comparison with each other, you must, of course, have a "feel" for the kinds of responses and range of writing the exercise elicited. In the training part of the scoring session, the Chief Reader will present you with test papers selected to give you an idea of the types of papers you will be scoring. As you and the other readers score and discuss these training papers, you will consider questions that pertain to the exercise you are scoring. For example, the questions to be resolved for one exercise might be: How closely should the directions be followed? What is considered off the topic? How important are spelling, punctuation, and grammar? The ways that the group agrees to answer these questions become your standards for scoring that exercise. Standards are not prescribed by the Chief Reader; rather, they are set during the training part of the scoring session. As the nature of the exercise and the types of student responses vary from one test administration to the next, so will the standards vary.

THE TRAINING SESSION

1. Understand the Four-Point Rating Scale. You will score the test papers on a four-point scale with 4 as the highest score. Obviously, there is no middle score. One of the reasons why the four-point scale works well is that readers are forced away from a middle, or uncommitted, score. You must first decide whether a test paper, when compared with the other test papers, is in the upper half (the 4-3 category) or in the lower

half (the 2-1 category). Then you must decide whether that paper is written well enough to rate the highest score (4) or so poorly that it rates the lowest score (1). New readers are often reluctant to grant any paper a 4. You must remember that the papers are being judged against each other, not against a "perfect" paper that might exist in your mind. Unfortunately, no reader — experienced or inexperienced — seems to need assurance about giving out 2's and 1's; all readers need to be reminded from time to time that some papers are worth more than a 2 or a 1.

2. Score the Training Papers. The Chief Reader and the Assistant Chief Reader will select test papers that will give you an idea of the types of responses to expect. You will read photocopies of the training papers and assign them scores. You may have doubts about how to score the first training papers, but after you have read six or seven, you should be able to compare papers and make judgements readily.

As you score the training papers, the Chief

Reader will record the score distribution on a chalkboard. If, after scoring several training papers, you find that your scoring is inconsistent with that of the rest of the readers, you should adjust to the standards of the majority. Unless all readers judge the papers in the same way, the scoring procedure will be unreliable and unfair to the students.

Discussing Problems and Unusual Papers

During the reading, you might receive a test paper that is difficult to score because it is off the topic, illegible, or upsetting. You should always take such a paper immediately to the Chief Reader for consultation. Unusually creative papers — those that are really poems, short stories, or plays — should also be taken immediately to the Chief Reader. These are special papers, and they must be scored in a special way. In the training session, readers should set a policy for handling these papers. (If a more imaginative paper relates at all to the task assigned, you would probably be wise not to downgrade the student for allowing his or her imagination free rein. Just be certain that you never lose sight of the standards set by you and the other readers.)

Papers whose authors have made little or no attempt to do the exercise (blank papers, papers with only a few ineffective words, or papers that merely repeat the assignment) should receive a score of 0. These must be taken to the Chief Reader as well.

The Scoring

Aides continually distribute and collect the test papers. You are to read quickly and make judgements based on your first impression of that piece of writing. The Chief Reader will monitor the scoring by reading various test papers after they have been scored. If the Chief Reader finds that several of you are losing sight of the standards or that there are too many score discrepancies, he or she may interrupt the scoring so that you may all be brought back into agreement.

Because each test paper being scored holistically is assessed by two different readers, it will have two scores on it. If both scores match, the readers were in perfect agreement. If the score combination is a split of either 4/3, 3/2, or 2/1, this split score is acceptable. However, discrepant scores from readers (combinations of either 4/2, 4/1, or 3/1) are not acceptable and must be resolved. Test papers with discrepant scores are read and scored again by two different readers. If these readings also result in discrepant scores, the Chief Reader and the Assistant Chief Reader must assign the paper its final scores. All scores are reported as the sum of the final two scores.

4. FOR THE CHIEF READER AND THE ASSISTANT CHIEF READER

Your scoring session will be successful if you become thoroughly familiar with the critical components of a successful reading. These are:

- A thorough understanding of the holistic method, including how scoring standards differ from traditional criteria for grading, the method's legitimate applications as well as its limitations, and differences between scoring standards and minimal standards;

> Section 1 of this manual covers these points.

- The selection of representative training papers that exemplify hypothetical score categories, and illustrate diverse approaches to the writing assignment;

> This section of the manual covers the selection of training papers in depth.

- Setting scoring standards that reflect the levels of writing to be assessed, accommodate readers' diverse values, yet yield sufficient agreement to ensure the consistency of scores;

> Setting scoring standards is discussed in this section of the manual and in section 3, "For the Readers."

- Thorough and efficient procedures for planning and conducting readings so that participants find scoring sessions professionally rewarding rather than burdensome and unproductive.

> Step by step procedures follow. You will also want to read section 2, "For the Administrator" and section 5, "For the Aides."

Several Weeks Before the Reading

selecting and scoring training papers

WHAT ARE TRAINING PAPERS?

Training papers, or range finders, are sample test papers selected to give readers an idea of the types of student responses to expect. Training papers illustrate both the levels of writing (represented by the score scale 4 to 1), and the different approaches students took in addressing the topic. Before scoring begins, the Chief Reader introduces training papers for purposes of setting standards (deciding what types of writing will be scored 4, 3, and so on), and acquainting readers with problems they may encounter (for example, a paper that is well-written but off topic).

> Choosing the training papers is a crucial step in the holistic scoring process. It is only by discussing the training papers that the readers arrive at their group standards or guidelines for scoring the students' responses to a particular exercise.

WHEN SHOULD TRAINING PAPERS BE SELECTED?

As soon after the test administration as possible, read enough student papers for each exercise to get a "feel" for the different ways students have responded to the assignment. Because training papers must be prepared for the scoring session, allow sufficient lead time, ideally <u>one month</u>, but at least <u>two weeks</u>.

HOW MANY PAPERS SHOULD YOU READ?

You should read at least 35 - 50 student responses to each exercise or writing assignment. These papers should be selected at random from different classes in your school or district.

HOW MANY AND WHAT TYPES OF PAPERS SHOULD YOU SELECT?

For each exercise, select 15 - 20 papers that you think exemplify the various types of student responses you have read.

Select papers that represent, as a group, the full range of the student writing you found for each exercise. There should be some best (score of 4) and worst (score of 1) papers for each exercise, but most should be examples of the average papers (scores of 3 and 2) written for this administration. When all of the test papers have been scored, there will probably be about as many 4 papers as there are 1 papers, as many 3's as there are 2's. As you would expect, there will be many more 3's and 2's than 4's and 1's. The range of writing in the 3 and 2 categories is always much wider than in the 4 and 1 categories.

Avoid using as a training paper the very best or the very worst paper you find as representing a typical 4 or 1 paper. For every score category, there is a range of appropriate responses; readers can best realize this range if they see an example of a middle-4 paper and a middle-1 paper. In the actual scoring, they will also give 4's and 1's to some papers that are slightly better and to some that are slightly poorer than the 4 and 1 training papers.

Select some problem papers. Some of the papers you choose should be ones that readers will have difficulty scoring, either because it appears that the writer did not follow the directions exactly or because the writer responded in an unusual way. For example, if the directions for the exercise say, "Give one reason for...," select at least one test paper that gives several reasons. In the training part of the scoring session, the readers will decide how to judge such a paper.

If the directions say, "Decide which proposed use of the building you support," select a paper that effectively rejects all of the proposed uses and supports none. Again, it will be up to the readers to determine how to judge such a response.

If the directions say, "Write a letter...," select papers that are written effectively but are not in letter format. Similarly, select other papers that are in perfect letter format but that are weak in content.

Select also other papers that are unusually creative, especially if they are only marginally related to the assignment.

All 15-20 papers that you select for the first exercise to be scored will become the basic training papers for the holistic scoring session.

HOW SHOULD YOU SCORE THE TRAINING PAPERS?

Using the score scale 4 (best) to 1, the Chief Reader and Assistant Chief Reader should score the papers independently.

DO NOT write scores or comments on the papers because these papers will be photocopied for the readers.

When you have both scored the papers, discuss your ratings, paper by paper. If you cannot resolve a score difference, replace the troublesome paper with one you both agree should be scored a 4, 3, 2, or 1.

HOW SHOULD YOU ORDER THE PAPERS FOR PRESENTATION?

Place the training papers in the order in which you will present them to the readers. First, select six papers that you both think clearly represent scores 4, 3, 3, 2, 2, and 1 for this test administration. From these six, choose either a middle-3 or a middle-2 paper as your first training paper. Then mix the remaining five papers. (See Example of Master List on the following page.)

Put the rest of the training papers in the order in which you want to present them. Remember to keep scores mixed.

HOW SHOULD YOU LABEL THE TRAINING PAPERS?

Label all of the training papers (in the order in which you decide to present them) by writing an "A," "B," "C," and so on in a large block letter in the upper right-hand corner of the paper. If you think that teachers may interpret these letters as designating grades, begin labeling with "G," "H," "I," and so on.

Since you will not be able to write on students' papers,

you will need to keep track of the scores assigned to each training paper. Therefore, make a master list.

Example of Master List

Preliminary Scores

Training Papers

Paper	Score
A	2
B	1
C	3
D	2
E	4
F	3

WHAT FINAL PREPARATIONS ARE NEEDED?

1. Check all of the training papers to be sure they are dark enough to photocopy well. If a paper is too light, copy it and go over the writing on the copy before proceeding further.

2. Have training papers photocopied, one copy for each reader, a copy each for both of you.

3. Keep all "A" papers in a stack, all "B" papers in a stack, and so on, so that aides can distribute <u>one</u> training paper at a time.

4. Return the original copy of each training paper to the batch to be scored. Mix them in with other papers.

Allow enough time between the selection of training papers and the scoring session to make the necessary copies of the training papers for each exercise.

making room arrangements

WHAT IS THE MOST EFFICIENT WAY TO ARRANGE THE ROOM FOR THE SCORING SESSION?

Specific room arrangements will depend, of course, on the facilities you have and the number of readers you must accommodate. Guidelines for determining number of readers are presented in section 2 of this manual,"For the Administrator." Once the size of the reading is known, you will want to work with your administrators in planning room arrangements.

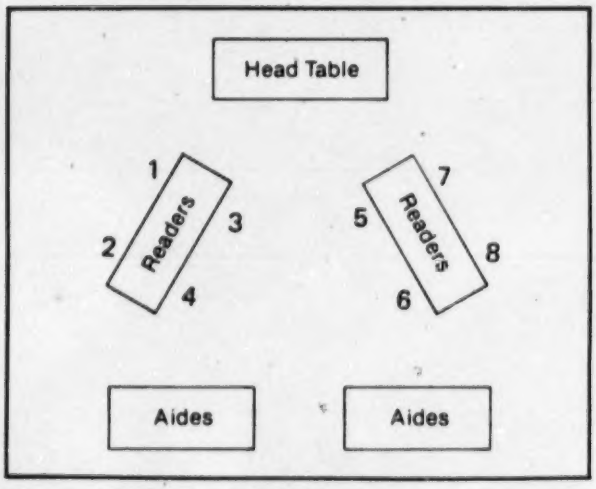
Three sample room arrangements are presented on page 31. Whichever plan is chosen or designed, several requirements must be kept in mind:

- The room must be well-lighted.
- The room must be large enough to accommodate the necessary tables, allowing aides to circulate without disturbing readers.
- Readers' tables must not be crowded; each reader should be comfortable and have sufficient space for papers to his or her right and left.

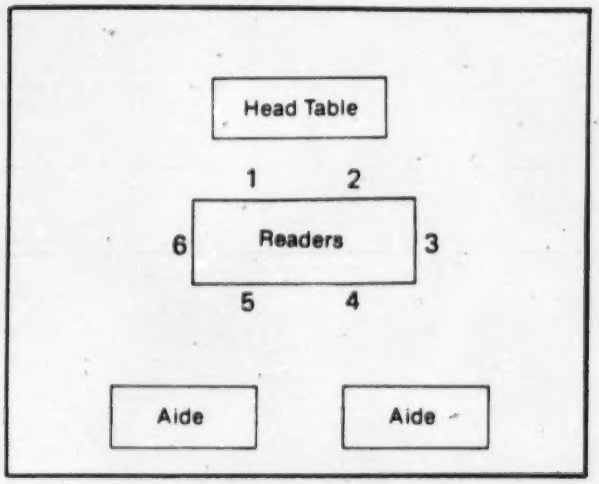
The following equipment will be needed:

- A chalkboard
- Tables to seat readers
- A head table for the Chief Reader and Assistant Chief Reader
- Separate tables for aides distributing papers and aides checking scores.

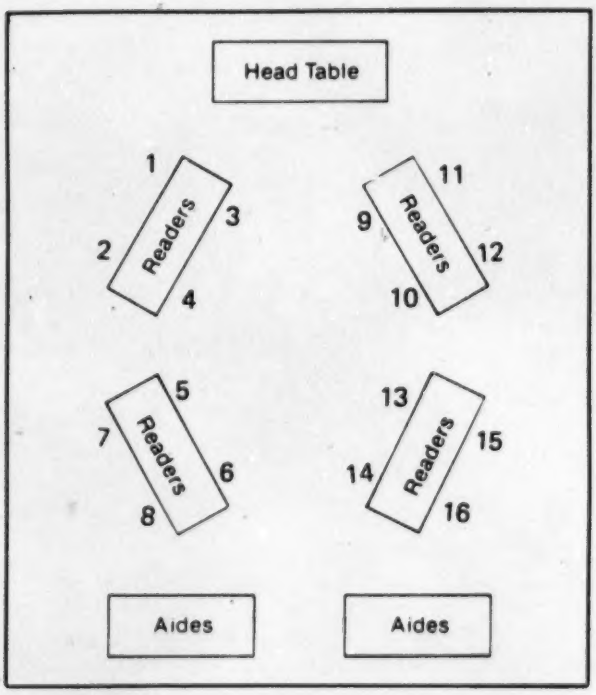
Arrangement of Tables at Scoring Session



8 Readers (Average size reading)



4-6 Readers (Small reading)



16 Readers (Large reading)

scheduling the scoring session

The length of a reading varies with the number of readers, the number of papers to be read, the length of the exercise the students worked on, and the experience of the Chief Reader and the readers. These factors must be determined before a schedule is planned. (See section 2, "For the Administrator," and the Exercise, "Planning the Length of the Scoring Session and Estimating the Number of Readers Needed.")

A sample schedule is presented on page 33 . This agenda, for half a school day, assumes 10 readers, 300 student papers for two exercises (each read twice = 1,200 readings). The schedule is given as an illustration rather than suggested as a model. Important things to note are the time allocated to training, the number of breaks scheduled, and the amount of time available for scoring (compared to the time consumed by training, breaks, and lunch).

You can safely assume that it will probably take 30 - 40 minutes to train readers for the first exercise, with less time (approximately 20 - 25 minutes) for subsequent exercises.

Readers easily tire of sitting and reading, so you should plan a stand-and-stretch break every half hour and at least a 10-minute coffee break in the morning and in the afternoon. If the scoring of an exercise is interrupted for lunch or some other comparable length of time, you should, before resuming the scoring,

present three or four new training papers to assure that the group is back on the right (that is, the agreed-upon) track.

After you have looked over the sample agenda, you should meet with administrators to discuss: the length of the scoring session, the number of readers needed, and when the reading is to take place (that is, a full school day, two afternoon sessions, and so on).

Sample Agenda

(10 Readers Scoring 300 Papers for One Exercise)

12:30 - 1:10	Introduce the first exercise and train the readers
1:10	Stand and stretch
1:10 - 1:40	Score papers
1:40 - 1:50	Coffee break
1:50 - 2:00	Present two or three additional training papers for the readers to score together
2:00 - 2:40	Continue scoring papers
2:40	Stand and stretch
2:40 - 3:20	Finish scoring papers and discuss results

planning the distribution system for the scoring session

HOW SHOULD YOU ORGANIZE THE DISTRIBUTION OF STUDENT PAPERS TO READERS?

Determine a method of having aides mix the test papers to be scored so that the papers from any one English class (or from any one school, if this is a district-wide scoring session) are not together. You will find it easier to control distribution if the papers are divided into groups of 15, 20, or 25 and placed into folders. At the scoring session, you may want to place the readers in two groups, one on each side of the room, so that papers scored on one side can simply be given to readers on the other side for a second scoring. You will have to devise a distribution system that best suits your situation, and you will have to make sure that the aides understand it.

devising a scoring sheet

WHY SHOULD YOU HAVE A SCORING SHEET?

In order to prevent readers from influencing each others' scoring, a procedure has been established to record scores on separate forms. Each folder of papers has two identical forms - one for each reader to use in recording scores. A set of sample forms might look like the example on the following page.

Each paper in the folder is given a number. The reader records his or her scores directly on the form, <u>not on the actual papers</u>. After the first reader has scored all the papers in a folder an aide collects the folder, removes the first scoring sheet and distributes the folder

Example: Scoring Forms

Folder #7	
Paper Number	Score
1	
2	
3	
4	
.	
.	
.	
20	
First Reader's Signature:	

Folder #7	
Paper Number	Score
1	
2	
3	
4	
.	
.	
.	
20	
Second Reader's Signature:	

to the second reader. When the second reader has finished scoring the set of papers, an aide can easily match the two scoring sheets and circle discrepant scores. Discrepant scores should be brought to your attention.

Two Days Before the Reading

WHAT SHOULD YOU DO?

1. Review the exercises and the training papers.

2. Read carefully the information on conducting the reading.

3. Meet with aides to be sure they understand what they are to do.

4. Use the quality control checklist at the end of Section 5 of this manual to ensure that all materials and arrangements are in order.

Conducting The Reading

You will need at your table

- copies of this part of the manual
- copies of training papers in the order you will present them
- copies of each exercise to be scored
- an agenda
- pencils, paper, chalk, eraser

preliminary discussion with readers

1. Assemble readers.

2. Explain the purpose of the reading, noting the number of papers to be scored.

3. Explain holistic scoring, emphasizing the fact that the guidelines for scoring the test papers will come from the group itself and that the papers will be judged against the other papers from this administration, not against student writing — either real or imagined — that was not produced for this test administration. (See "For the Readers".)

4. Explain the scoring procedure and the four-point scale. (See For The Readers.) If you are setting minimal standards, tell the readers that they will establish a minimally acceptable level of writing for this exercise after they have seen a sampling of the papers for it.

training the readers

1. Have aides distribute to readers the photocopies of the exercise (test questions and directions) about to be scored.

2. Read the directions aloud, stressing what the student was to do. If there are questions, keep the discussion pertinent. Mention any major problems that you have detected in the students' responses.

3. Have aides distribute photocopies of training paper A.

4. Tell readers to read quickly and consider whether the paper will probably be in the upper or lower half of the full range of

papers, then consider how high (4 or 3) or how low (2 or 1) it will fall in that half. Have readers write their scores on the training paper. Of course, readers may be uneasy about trying to pinpoint a score at this early stage. You may need to reassure them that they should not expect agreement yet, and that they will reevaluate this paper after they have read other training papers.

5. Ask the readers: "How many gave the paper a score of 4? A score of 3? Of 2? Of 1?" Then tally the scores on the chalkboard. The distribution of these early scores might look like this:

Readers' Scores for Training Papers

	(Score) 4	(Score) 3	(Score) 2	(Score) 1
Paper A	—	I	II	~~IIII~~
Paper B				
Paper C				
Paper D				
Paper E				
Paper F				

6. Repeat the process (distribute, score tally) for the next five training papers (B to F). The scores should cluster, indicating group consensus. If the scores for any paper are spread across three or all four points of the scale, call for

a discussion. Ask readers to explain why they scored it the way they did. The paper obviously presents them with a problem that needs to be resolved. By discussing problems like these, readers should be able to establish their guidelines for this scoring session.

7. Have readers order these six training papers (A through F) from best to worst. This step reinforces to the readers that each paper is to be judged in comparison with the others. Allow no talking; judgements must be independent, even in the training session.

8. Ask the readers for the order in which they ranked the papers. For a small group of readers, you might want to tally the ranking on the chalkboard to show them the extent of agreement. The tally could look something like this:

Rank Order of the First Six Training Papers

	Best	2nd	3rd	4th	5th	Worst
Paper A				𝍸	III	
Paper B						𝍸 III
Paper C		I	𝍸 II			
Paper D				III	𝍸	
Paper E	𝍸 III					
Paper F		𝍸 II	I			

For this hypothetical group of eight readers, you can see that there was complete agreement about the best and worst papers. There was minor disagreement about the placement of the middle papers. This is to be expected. For every point on the scale, there is a range of papers and the range is greater for the 3 and 2 points. That is, there will probably be a variety of student writing receiving either a 4 or a 1, but a greater variety of student writing will receive either a 3 or a 2.

Because there is a range of student writing within each score category, some of the weakest 4 papers will not be much better than some of the best 3 papers. Some of the lowest 3 papers may not be much better written than the best of the 2 papers, and so on down the scale. These end-of-the-range papers may very well receive two different scores from two different readers. If the two scores are separated by only a one-point difference (4/3,3/2, or 2/1), they are called "split scores." A split score is not only acceptable but is, for many papers, the most accurate score. When the two scores are added, the 3/2 split, for example, will result in a final score of 5. That paper has been judged better than the 2/2 papers (with a final score of 4) and not as good as the 3/3 papers (with a final score of 6).

9. Have aides distribute the next training paper (G). Repeat the process (read quickly, mark scores, tally). Continue with several more training papers, one at a time.

10. By now, readers should have reached consensus. If one or two are not scoring with the majority, you will have to identify the reasons for this and resolve the problem. The success of holistic scoring depends on the degree of inter-reader agreement. If you think the group needs more training, present a few more training papers.

11. If you are setting minimal standards in this scoring session, ask the readers to identify training papers that they consider minimally acceptable in both content and expression. Explain that this minimally acceptable level of writing is assigned a score on the scoring scale. Have the readers decide on the minimal score. For example:

Score

4 - Is superior

3 - Meets minimal local standards

2 - Does not meet minimal local standards

1 - Is extremely weak

You may need to stress to the readers that not all of the papers assigned the minimal score will precisely match the minimally acceptable papers but that the overall writing ability should be similar. (Note: The distribution of scores will not necessarily fall into a normal curve when you are setting standards.)

12. When the readers have resolved all of the questions that you thought should be raised about scoring the test papers for this exercise, put the remaining training papers aside so that you can use them if further training is necessary later in the scoring session.

13. Discuss how to handle unusually emotional or unusually creative papers that are difficult to score even with the help of the quidelines. Tell the readers to bring these papers immediately to the head table for consultation. (See "For the Readers.")

scoring the papers

Throughout the reading you should

- Allow no unnecessary talking among readers or aides.
- See that the papers are distributed as planned.
- Monitor the readers by reading several of the papers they have scored. If you find that a reader is not scoring according to the guidelines established in the training period, casually discuss the problem with that reader. If several readers are scoring out of line, you may need to retrain the group with a few more training papers.
- Resolve scores that are discrepant.
- Recognize readers' fatigue and call for unscheduled breaks if needed.

1. Have the aides distribute the papers to be scored for this exercise.

2. When all of the papers for an exercise have been scored two times and all discrepant scores resolved, the reading for that exercise is over.

the next exercise

For the next exercise to be scored holistically, the training process may be considerably shorter. You will, however, follow the same procedures of:

1. Discussing the topic.

2. Reading and scoring the first six training papers.

3. Establishing guidelines and possibly identifying minimally acceptable papers for setting standards.

4. Reading a few more training papers to reach scoring consensus.

5. Scoring the actual test papers.

5. FOR THE AIDES

Responsibilities

You are to assist the Chief Reader by preparing materials for the exercises to be scored. You will participate in the scoring session by distributing papers, recording scores, and following the specific instructions of the Chief Reader.

The Chief Reader may ask you to fold over and staple the cover of the test booklets so that the readers will not see the students' names. The Chief Reader may also ask you to separate test booklets so that several exercises can be scored at the same time by different groups.

several weeks before the scoring session

1. Review all your responsibilities and meet with the Chief Reader to discuss them.

2. Make photocopies of each training paper given to you by the Chief Reader. The Chief Reader, the Assistant Chief Reader, and all of the readers will need their own copies of each training paper.

3. Make photocopies of the exercise (test question and directions) to be scored. (Make the same number of photocopies as you did in Step 2.)

4. For each exercise, mix students' test papers so that no reader has too many test papers from any one class. Put the papers in folders in groups of 25, 20, or 15 in order to control their distribution more easily. Number the papers in each folder; for example 1 to 20.

5. Make scoring sheets. You will need two scoring sheets for each folder of papers. These two forms will be identical except that one will specify "First Reader's Signature" and the other, "Second Reader's Signature". Be sure that the paper numbers on each form correspond to the actual number of papers in each folder. In other words, if you have 20 papers in each folder, there must be 20 numbers listed on each scoring sheet.

Folder #7	
Paper Number	Score
1	
2	
3	
4	
.	
.	
.	
20	
First Reader's Signature:	

Folder #7	
Paper Number	Score
1	
2	
3	
4	
.	
.	
.	
20	
Second Reader's Signature:	

6. Put copies of the two companion scoring sheets in each folder.

the day before the scoring session

1. See that the room is set up according to the Chief Reader's instructions.
2. Put at each reader's place and at the head table:
   - A supply of lined paper
   - Two sharpened pencils
3. Be sure that the aides' tables have:
   - Photocopies of the exercise to be scored
   - Photocopies of the training papers
   - All student papers for the exercise that is about to be scored
4. Arrange students' test papers for distribution according to directions from the Chief Reader.

during the scoring session

When readers are trained

1. Distribute photocopies of the exercise (test questions and directions) when the Chief Reader tells you.
2. Distribute training papers when the Chief Reader tells you.

When scoring begins

1. According to instructions from the Chief Reader, distribute the papers to be scored.

2. Always place test papers on a reader's left and remove from a reader's right.

3. Distribute and collect papers continuously until all have been read twice. Do not allow a reader to sit idly waiting for more papers.

4. After a paper has been read twice, it goes to the recording aide who checks the scores. If the two scores are discrepant (such as 4/2, 4/1, or 3/1), the recording aide should remove the paper from the batch that has been scored satisfactorily. Then the recording aide puts that test paper back into the distribution scheme so that it can have a third and fourth reading by different readers. If the two new scores are also discrepant, the test paper goes to the Chief Reader for resolution. Keep the Chief Reader informed of the number of discrepant scores.

5. If a reader hands you a test paper that has a score of zero, do not give that paper to a second reader but take it immediately to the Chief Reader.

6. When most of the papers have been read, give readers only two or three papers at a time to ensure that all readers finish at about the same time.

Chief Reader and Aides'

Quality Control Checklist

Room Arrangements

Head table ____

Readers' tables ____

Aides' table(s) ____

Chalkboard ____

Training Papers

Photocopied in sufficient number ____

Stacked in the order you will present them ____

Master list of scores assigned to training papers ____

Student Papers

Training papers returned to batch ____

Mixed rather than grouped by class or district ____

Organized for quick distribution ____

Chief Reader's Quality Control Checklist (cont.)

Materials for Each Reader

A copy of each writing exercise _____

A supply of lined paper ____

Two sharpened pencils ____

Removable tape or a card to cover discrepant scores ____

Arrangements for Coffee ____

DEFINITIONS

Chief Reader .....Person who organizes and conducts the scoring session; person who is trained in holistic scoring.

Discrepant scores.....Scores that are separated, by more than 1 point (e.g. 4/1, 3/1, 2/4); If two readers score a paper discrepantly, the paper must be scored again by two other readers.

Holistic Scoring.....Judging a piece of writing for its total effect rather than assessing each element (e.g., organization, mechanics) separately.

Minimally acceptable score.....Point on the score scale at or below which a paper will be classified as not meeting basic district standards. This is also referred to as a cutting score.

Problem papers.....Papers which, for one reason or another, are very difficult to score. For example, a paper that is especially well written but off the topic.

Reading.....A session during which readers establish guidelines and score papers.

Score scale.....Point range, usually from 4 (best) to 1, used in rating papers.

Scoring standards.....Guidelines or criteria that readers establish for rating papers either 4,3,2 or 1. These guidelines are not to be confused with minimal standards.

Split scores.....Scores that are separated by 1 point (e.g., 4/3, 3/2). Split scores often represent the best assessment of a paper. They do not need to be resolved.

Training papers.....Papers selected by the Chief Reader to represent the range of writing on a given topic. Training papers (also called range finders) are used to establish scoring standards and to train readers to score consistently.